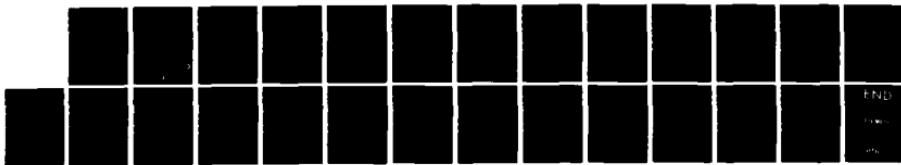


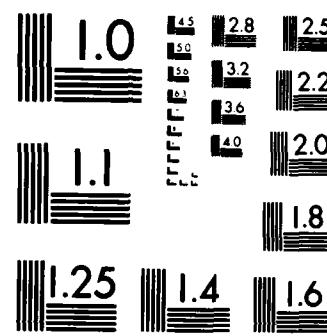
AD-A150 096 A DIFFUSION PROCESS AND ITS APPLICATION TO DETECTING A 1/1
CHANGE IN THE DRIFT OF BROWNIAN MOTION(U) STANFORD UNIV
CA DEPT OF STATISTICS M POLLAK ET AL. SEP 84 TR-30

UNCLASSIFIED N00014-77-C-0306

F/G 7/4

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AD-A150 096

(1)

A DIFFUSION PROCESS AND ITS APPLICATION TO DETECTING
A CHANGE IN THE DRIFT OF BROWNIAN MOTION

by

Moshe Pollak
Hebrew University of Jerusalem

and

David Siegmund
Stanford University

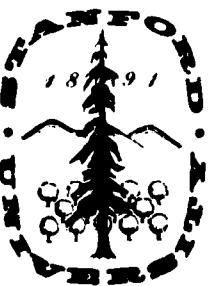
TECHNICAL REPORT NO. 30
SEPTEMBER 1984

PREPARED UNDER CONTRACT
N00014-77-C-0306 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



DTIC
ELECTED
FEB 11 1985
S B D

DTIC FILE COPY

84 11 14 174

DISCLAIMER NOTICE

**THIS DOCUMENT IS BEST QUALITY
PRACTICABLE. THE COPY FURNISHED
TO DTIC CONTAINED A SIGNIFICANT
NUMBER OF PAGES WHICH DO NOT
REPRODUCE LEGIBLY.**

**A DIFFUSION PROCESS AND ITS APPLICATION TO DETECTING
A CHANGE IN THE DRIFT OF BROWNIAN MOTION**

by

**Moshe Pollak
Hebrew University of Jerusalem**

and

**David Siegmund
Stanford University**

**TECHNICAL REPORT NO. 30
SEPTEMBER 1984**

**PREPARED UNDER CONTRACT
N00014-77-C-0306 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH**

**Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government**

Approved for public release; distribution unlimited

**Also prepared under National Science Foundation Grant MCS80-24649, and issued as
Technical Report No. 222, Department of Statistics, Stanford University.**

**DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA**

A Diffusion Process and Its Application to Detecting a Change in the Drift of Brownian Motion

Moshe Pollak

Hebrew University of Jerusalem

and

David Siegmund
Stanford University

1. Introduction.

Consider a Brownian motion process $W(t)$, $0 \leq t < \infty$, which during the time interval $[0, \nu]$ has drift 0 and during (ν, ∞) has drift $\mu > 0$, where $\nu \leq \infty$ and μ are unknown parameters. We seek a stopping rule T which "detects" the change point ν "as soon as possible." For example, $W(t)$ may represent the cumulative output of an industrial process, which is under control so long as the average output is 0, but which may go out of control and then must be corrected as soon as possible. Other domains of application are to maintaining quality of repeated assays (Wilson, et al., 1970) and surveillance of birth records for a possible increase of genetic malformations (e.g. Weatherall and Haskey, 1976).

Let P_ν denote probability when the change occurs at time ν ($\nu \leq \infty$). (The dependence on μ is suppressed. When it seems desirable to emphasize this dependence, we shall write $P_{\nu, \mu}$. Note that $P_\infty = P_{0,0}$.) A stopping rule T to detect the change point should have a large value for $E_\infty(T)$, i.e. if no change occurs, the expected time until one is "detected" should be large. Subject to $E_\infty(T)$ being large, a good detection rule should in some sense have small values of $E_\nu(T - \nu \mid T > \nu)$, i.e. the time after a change occurs until it is detected should be small. Some common detection rules (including those considered in this paper) satisfy $\sup_\nu E_\nu(T - \nu \mid T > \nu) = E_0(T)$, in which case a detection rule can to some extent be evaluated in terms of its Average Run Lengths: $E_\infty(T)$, which should be large, and $E_0(T)$, which should be small.

One possible solution to the detection problem is given by the so-called cusum tests proposed by Page (1954). For a systematic discussion of these procedures, see van Dobben de Bruyn (1968). An outstanding contribution to the substantial literature on cusum processes is Lorden (1971), who shows that they are asymptotically optimal when $E_\infty(T)$ is infinitely large.

Shiryayev (1963) and Roberts (1966) independently proposed the same competitor to cusum tests. Recently Pollak (1984) has proven an optimality property for the Shiryayev-Roberts rule (in discrete time – see Appendix A for a brief discussion of this result in the present setting) which seems considerably stronger than Lorden's asymptotic optimality of cusum stopping rules.

The purpose of the present paper is to make a quantitative comparison of the Shiryayev-Roberts and Page procedures. We do this in the context of continuous time in order to use the machinery of diffusion processes to perform explicitly certain calculations, which seem impossible in discrete time. (See, however, Pollak, 1983, who makes considerable progress on the evaluation of average run lengths in discrete time.) Although the continuous time results are not especially good approximations to the corresponding quantities in discrete time, they provide very useful comparative information on which to base selection of a stopping rule.

The paper is organized as follows. The Shiryayev-Roberts process is defined in Section 2, and shown to be a novel diffusion process with some surprising properties. We also specify more precisely the basis for our comparison of the two procedures and give the results of some elementary calculations. These developments continue in Section 3, which contains an asymptotic evaluation of $E_\nu(T - \nu \mid T > \nu)$ when ν and $E_\infty(T)$ are large. In Section 4 we define a modification of our basic procedure and give an asymptotic evaluation of its average run length. Numerical comparisons and a discussion of their significance are contained in Section 5. Some open problems are mentioned briefly in Section 6. The reader whose principal interest is in our conclusions may wish to read Section 2 (through Proposition 1) and then skip directly to Sections 5 and 6, before returning to the derivations. Our conclusions are roughly these. In simple situations where the two procedures can be directly compared,

neither seems dramatically superior to the other. However, the Shirayev-Roberts procedure is more easily adapted to complex circumstances and consequently warrants additional study. (See Section 6 for examples.)

2. Definition of the Procedures, Criteria for Comparison, and Elementary Operating Characteristics.

Suppose momentarily that at time ν the drift of $W(t)$ changes from 0 to some known value $\delta > 0$. Although this will rarely be true, it is possible that a procedure derived under this hypothesis is useful even if the drift μ after time ν is unknown. Also, for the sake of motivating our stopping rule, suppose that ν is itself a random variable which is exponentially distributed with mean $1/\lambda$. Then the posterior probability that there has been a change before time t given the data until t is

$$P\{\nu < t \mid W(s), s < t\} = \frac{\int_0^t \exp(-\lambda s) \exp[\delta\{W(t) - W(s)\} - \delta^2(t-s)/2] ds}{\int_0^t \lambda \exp(-\lambda s) \exp[\delta\{W(t) - W(s)\} - \delta^2(t-s)/2] ds + \exp(-\lambda t)}.$$

Consider the rule which stops and declares that a change has taken place when this posterior probability exceeds some threshold c for the first time. (For a particular loss structure Shirayev, 1963, has shown that the Bayes rule has this form.) For λ close to 0 this stopping rule is approximately

$$T = T_B = \inf\{t : \int_0^t \exp[\delta\{W(t) - W(s)\} - \delta^2(t-s)/2] ds \geq B\}, \quad (1)$$

which is a stopping rule proposed by Shirayev (1963) and Roberts (1966).

Page's rule is similar, but is motivated by maximum likelihood rather than Bayesian considerations. It is defined by stopping at

$$\hat{T} = \inf\{t : \delta[W(t) - \delta t/2 - \min\{W(s) - \delta s/2\}] \geq c\}. \quad (2)$$

In principle we would like to choose B and c , so that $E_\infty(T) = E_\infty(\hat{T})$, then compare $E_\nu(T - \nu \mid T > \nu)$ and $E_\nu(\hat{T} - \nu \mid \hat{T} > \nu)$ as functions of both ν and μ . In fact our comparisons are basically between the extreme cases $\nu = 0$ and $\nu = \infty$. Sometimes they are asymptotic as B (hence also c) $\rightarrow \infty$. The easiest comparisons are, of course, when we suppose that the only possible value of μ is the hypothesized value $\mu = \delta$; but we shall also

consider other possible values. (In Section 4 we introduce a modification of the stopping rule (1) which is designed to deal with the case of unknown μ .)

We begin with a detailed examination of the stopping rule (1). Let

$$R(t) = \int_0^t \exp[\delta\{W(s) - W(0)\} - \delta^2(s-t)/2] ds. \quad (3)$$

Since for fixed s $\exp[\delta\{W(t) - W(s)\} - \delta^2(t-s)/2]$ is a P_∞ -martingale in t for $t \geq s$, it follows that $R(t) - t$ is also a P_∞ -martingale. Also $E_\infty\{R(t)\} = t$. An easy application of optional stopping (Loeve, 1963, p. 534) yields

Proposition 1. $E_\infty(T) = E_\infty\{R(T)\} = B$.

Remark: The martingale property of $R(t)$, which yields a simple formula for $E_\infty(T)$, proves to be very useful in adapting the Shirayev-Roberts rule (1) to deal with more complicated problems. See Section 6 for some examples.

From (3) it is easy to see that $R(t)$ is a Markov process with stationary transition probabilities (as long as the drift of $W(t)$ does not change). It follows from Ito's formula that for all μ the $P_{0,\mu}$ stochastic differential of $R(t)$ is given by

$$dR(t) = \{1 + \mu\delta R(t)\}dt + \delta R(t)d\tilde{W}(t), \quad (4)$$

where $\tilde{W}(t)$, $0 \leq t < \infty$, is standard Brownian motion (with drift 0). Hence under $P_{0,\mu}$ the differential generator of $R(t)$ is given by

$$Df(x) = \frac{1}{2}\delta^2 x^2 f''(x) + (1 + \mu\delta x)f'(x) \quad (x > 0). \quad (5)$$

From (5) and standard diffusion theory it is possible to compute the average run lengths $E_{0,\mu}(T)$ in a fairly explicit form. A convenient reference for the following calculations is Karlin and Taylor (1981, Chapter 15).

The scale function $S_\mu(x)$ of the process $R(t)$ is determined by integrating the relation

$$S'_\mu(x) = \exp\{2/(\delta^2 x) - (2\mu \log x)/\delta\}, \quad (6)$$

and the speed measure is given by

$$dM_\mu(x) = dx/\delta^2 x^2 S'_\mu(x). \quad (7)$$

It is easy to see that

$$\int_0^1 \{S_\mu(1) - S_\mu(x)\} dM_\mu(x) < \infty \quad (8)$$

and hence 0 is an entrance boundary.

Since $R(t)$ is a Markov process with stationary transitions, we can consider the process starting from $R(0) = x$. When this is the case we shall write E^x and P^x to denote expectation and probability. For $a < x < b$ and $R(0) = x$, let $N = \inf\{t : R(t) \notin (a, b)\}$. Then for nonnegative functions h

$$E_{0,\mu}^x \left[\int_0^N h\{R(t)\} dt \right] = \int_a^b h(y) G(x, y; a, b) dM_\mu(y), \quad (9)$$

where

$$\begin{aligned} G(x, y; a, b) &= 2\{S_\mu(x) - S_\mu(a)\}\{S_\mu(b) - S_\mu(y)\}/\{S_\mu(b) - S_\mu(a)\} \quad (x \leq y) \\ &= G(y, x; a, b) \quad (x \geq y). \end{aligned}$$

Letting $a \rightarrow 0$, then $x \rightarrow 0$ and using (8), (9) yields the following result, obtained by Shirayev (1963) in the special case $\mu = \delta$.

Proposition 2. For the stopping rule T defined by (1) and for all μ

$$E_{0,\mu}(T) = 2 \int_0^B \{S_\mu(B) - S_\mu(y)\} dM_\mu(y), \quad (10)$$

where S and M are given by (6) and (7). In the special case $\mu = 0$ this becomes $E_{\infty}(T) = B$, in agreement with Proposition 1. In the special case $\mu = \delta$ (10) yields

$$E_{0,\delta}(T) = 2\delta^{-2} \{ \log A + \exp(1/A) \int_{A^{-1}}^{\infty} (\log z) \exp(-z) dz \}. \quad (11)$$

where $A = \delta^2 B/2$. Letting $B \rightarrow \infty$ yields

$$E_{0,\delta}(T) = 2\delta^{-2} \{ \log A - \gamma + O(A^{-1} \log A) \}, \quad (12)$$

where $\gamma \cong .5772$ is Euler's constant.

Remark: It is easy to evaluate (10) numerically. For some purposes one probably obtains more insight from a simple asymptotic expansion like (12). It is possible to obtain similar expansions when $\mu \neq \delta$, but the calculus and the resulting expressions are much more complicated. The details are omitted.

For the purpose of evaluating $E_\nu(T - \nu \mid T > \nu)$ as $\nu \rightarrow \infty$, it will be helpful to know the P_∞ limiting distribution of $R(t)$ (cf. Section 3). Since $E_\infty\{R(t)\} = t$, it is somewhat surprising that (under P_∞) $R(t)$ is actually recurrent.

Proposition 3. For any initial state $R(0) = x$, and for all $y > 0$

$$\lim_{t \rightarrow \infty} P_\infty^x\{R(t) < y\} = \exp(-2/\delta^2 y).$$

Proof. Let $\tau_z = \inf\{t : R(t) = z\}$. For fixed $u > x = R(0)$ let τ denote the time of first return to x after passing through u . Then

$$E_\infty^x(\tau) = E_\infty^x(\tau_u) + E_\infty^u(\tau_z),$$

and the right hand side may be evaluated by taking appropriate limits ($a \rightarrow 0$ or $b \rightarrow \infty$) in (9) with $h = 1$. The result is that

$$E_\infty^x(\tau) = 2\{S_0(u) - S_0(x)\}M_0(0, \infty), \quad (13)$$

which is finite by (6) and (7). Let $H(t) = P_\infty^x\{R(t) < y\}$.

By the standard renewal argument

$$H(t) = P_\infty^x\{\tau > t, R(t) < y\} + \int_0^t H(t-s)P_\infty^x\{\tau \in ds\}.$$

By (13) the renewal theorem applies to yield

$$\begin{aligned} \lim_{t \rightarrow \infty} H(t) &= \{E_\infty^x(\tau)\}^{-1} \int_0^\infty P_\infty^x\{\tau > t, R(t) < y\} dt \\ &= \{E_\infty^x(\tau)\}^{-1} E_\infty^x \left[\int_0^\tau I\{R(t) < y\} dt \right]. \end{aligned}$$

The numerator can be evaluated by the same limiting process that led to (13), now with $h(z) = I(z < y)$, to show that

$$\lim_{t \rightarrow \infty} H(t) = \int_0^y dM_0(z) / \int_0^\infty dM_0(z).$$

Using (6) and (7) to evaluate these integrals completes the proof.

The process $Y(t) = \delta[W(t) - \delta t/2 - \min\{W(s) - \delta s/2\}]$ which defines Page's stopping rule (2) is also a diffusion process, this time with a reflecting barrier at zero, so the same

theory delivers the average run lengths and P_∞ limiting distribution for this process as well. On the other hand, the relation of (2) to the stopping rule of a sequential probability ratio test, which was noted already by Page (1954), yields a more elementary computation of average run lengths; and the P_∞ limiting distribution is also easily computed by standard, direct arguments. We summarize the relevant results in the following proposition:

Proposition 4. Let \bar{T} be defined by (2). For all μ

$$E_{0,\mu}(\bar{T}) = 2(2\mu - \delta)^{-2} [c(2\mu - \delta)/\delta + \exp\{-c(2\mu - \delta)/\delta\} - 1], \quad (14)$$

where for $\mu = \delta/2$ the right hand side of (14) is defined to be $(c/\delta)^2$. For all $x > 0$, $y > 0$, as $t \rightarrow \infty$,

$$P_\infty^x\{Y(t) < y\} \rightarrow 1 - \exp(-y). \quad (15)$$

By Proposition 1 and (14), equating $E_\infty(T)$ with $E_\infty(\bar{T})$ means setting $A = \exp(-c) - c - 1$, where $A = \delta^2 B/2$. This can be asymptotically inverted as $A \rightarrow \infty$ to yield $c = \log(A) + \{\log(A) + 1\}/A + o(1/A)$. For the special case $\mu = \delta$, substitution of this relation into (14) yields

$$E_{0,\delta}(\bar{T}) = 2\delta^{-2} \{\log(A) - 1 + O(1/A)\}. \quad (16)$$

Comparing (12) and (16), we see that $E_{0,\delta}(\bar{T})$ is asymptotically smaller than $E_{0,\delta}(T)$, but the difference is not large enough to indicate a strong preference for Page's procedure. For $\mu \neq \delta$, the procedures are compared numerically in Section 5.

3. Asymptotic Evaluation of $E_\nu(T - \nu \mid T > \nu)$ as ν and $B \rightarrow \infty$.

In this section we try to compare T and \bar{T} when the time of change, ν , is large and the stopping rule has not yet signaled a change. The optimality considerations of Pollak (1984) (see also Appendix A) suggest that the Shirayev-Roberts rule should be better than Page's under these conditions. We shall see below that this expectation is essentially correct, but the difference is usually small.

A possible formulation is to evaluate $\lim_\nu E_\nu(T - \nu \mid T > \nu)$ with B fixed. However, this seems difficult technically and also inappropriate conceptually. In most applications we envision that the cost of a false alarm is substantial; and hence, at least insofar as we

are able to make crude prior judgments about the range of ν , we should choose B roughly comparable to ν - or perhaps larger. Here we shall suppose that ν and B are simultaneously large, which, by virtue of the following lemma, allows us to utilize the P_∞ unconditional limiting distribution of $R(s)$ calculated in Proposition 3.

Again let E^x (P^x) denote expectation (probability) when $R(0) = x$.

Lemma 1. For any $x, y, t > 0$

$$P_\infty^x\{R(t) < y \mid T > t\} \geq P_\infty^x\{R(t) < y\}.$$

Also $P_\infty\{R(t) < y\} \geq P_\infty\{R(\infty) < y\}$, where $R(\infty)$ denotes a random variable having the distribution evaluated in Proposition 3. As t and $B \rightarrow \infty$

$$P_\infty\{R(t) < y \mid T > t\} \rightarrow P_\infty\{R(\infty) < y\}.$$

Since this result seems potentially of wider interest than the specific technical requirements of the present paper, and since our proof uses essentially none of the structure of $R(t)$ beyond the fact that it is stochastically monotone, a complete proof of Lemma 1 will be published elsewhere.

Theorem 1. Let $A = \delta^2 B/2$. Suppose $\mu > 0$ and $B, \nu \rightarrow \infty$. For $\mu > \delta/2$

$$E_\nu(T - \nu \mid T > \nu) = \{\delta(\mu - \delta/2)\}^{-1}[\log A - \gamma - \delta/\{2(\mu - \delta/2)\} + o(1)].$$

For $\mu = \delta/2$

$$E_\nu(T - \nu \mid T > \nu) = \delta^{-2}\{(\log A - \gamma)^2 + \pi^2/6 + o(1)\}.$$

For $\mu < \delta/2$

$$E_\nu(T - \nu \mid T > \nu) = \{\delta(\delta/2 - \mu)\}^{-1}\{A^{1-2\mu/\delta}\Gamma(1 - 2\mu/\delta) - \log A + \gamma - (1 - 2\mu/\delta) + o(1)\}.$$

Here Γ denotes the gamma function and $\gamma = \Gamma'(1) \cong .5772$ is Euler's constant.

Proof. We start from

$$\begin{aligned} E_{\nu, \mu}(T - \nu \mid T > \nu) &= \int_0^B E_{0, \mu}^x(T_B) P_\infty\{R(\nu) \in dz \mid T > \nu\} \\ &= E_{0, \mu}^0(T_B) - \int_0^B E_{0, \mu}^0(T_z) P_\infty\{R(\nu) \in dz \mid T > \nu\}. \end{aligned}$$

To complete the proof, we first replace the measures $P_\infty\{R(\nu) \in dx \mid T > \nu\}$ by their limit, $P_\infty\{R(\infty) \in dx\}$ (cf. Lemma 1), and then evaluate the resulting integrals (cf. Proposition 3 and (6), (7)).

To justify replacing the distributions $P_\infty\{R(\nu) \in dx \mid T > \nu\}$ by their limit, it suffices to show that $g(x) = E_{0,\mu}^0(T_x)$ is uniformly integrable with respect to these distributions. But by Lemma 1, the distributions are stochastically smaller than their limit; and since g is monotone increasing, it is uniformly integrable if and only if it is integrable with respect to the limiting distribution. It follows from Proposition 2 and some calculation that for any $\mu > 0$ there exists $r < 1$ such that $g(x) = O(x^r)$. From Proposition 3 we see that $P_\infty\{R(\infty) > x\} \sim 2/\delta^2 x$, so g is in fact integrable. Evaluation of the resulting integrals to obtain the results stated above is sketched in Appendix B.

An essentially identical argument applies to Page's procedure. (In fact, as noted above, it is possible to abstract the essential features of Lemma 1 to cover both processes simultaneously.) We record the final result as Theorem 2.

Theorem 2. For T defined by (2), as c and $\nu \rightarrow \infty$, $E_\nu(\bar{T} - \nu \mid \bar{T} > \nu)$ is given approximately by the following expressions for the respective cases (i) $\mu \neq \delta/2$ and (ii) $\mu = \delta/2$:

$$(i) \quad \{\delta(\mu - \delta/2)\}^{-1}[c - 1 - \delta^2/\{2\mu(2\mu - \delta)\} + \{\delta/(2\mu - \delta)\} \exp\{c(1 - 2\mu/\delta)\} + o(1)],$$

and

$$(ii) \quad \delta^{-2}(c^2 - 2) + o(1).$$

If c is defined by the relation $A = \exp(c) - c - 1$, so that $E_\infty(T) = E_\infty(\bar{T})$, the results given in Theorem 2 are easily rewritten to be directly comparable to those of Theorem 1. For example, $E_\nu(T - \nu \mid T > \nu)$ is asymptotically smaller than $E_\nu(\bar{T} - \nu \mid \bar{T} > \nu)$ if $\mu < (1 - \gamma)^{-1}\delta = 1.13\delta$, but not otherwise. The differences are not large if $\mu > \delta/2$. These results explain the general conclusions of Roberts (1966), which in his case were based on a Monte Carlo experiment.

Remark: Shirayev (1963) considers the problem of "detection of destruction of a stationary regime," which has some technical points in common with the preceding discussion, but is conceptually quite different. See Appendix C.

4. Unknown μ .

In the preceding sections we have studied procedures which are approximately optimal under the assumption that at time ν the change from $\mu = 0$ is to a known value $\mu = \delta$. Since this assumption is never satisfied in practice, we have evaluated these procedures for general values of μ . Now we consider a generalization of the Shirayev-Roberts rule for the case of unknown μ .

Let G denote a probability on $(0, \infty)$. (The distribution G could be interpreted as a prior distribution for the value of μ after the change at time ν .) Let $R_\delta(t)$ be defined by (3), where now we use a subscript to denote dependence on the value of δ . Define

$$\hat{R}(t) = \int_0^\infty R_y(t) dG(y), \quad (17)$$

and let

$$\hat{T} = \hat{T}_B = \inf\{t : \hat{R}(t) \geq B\} \quad (18)$$

be defined as in (1), but with \hat{R} in place of R . Since $R_y(t) - t$ is a P_∞ martingale with mean equal to 0 for every value of y , it follows that $\hat{R}(t) - t$ is also a P_∞ martingale with mean 0. Hence, exactly as in Proposition 1

$$E_\infty(\hat{T}) = B. \quad (19)$$

Before turning to an evaluation of $E_{0,\mu}(\hat{T})$, we note that an analogous modification of Page's rule was defined and studied in Pollak and Siegmund (1975). Unfortunately, however, we do not know an approximation to the P_∞ average run length of this procedure because there is no similar martingale structure. Hence we were limited then, as we are now, to examining the $E_{0,\mu}$ average run length. In Section 3 we give other examples of processes for which a version of the Shirayev-Roberts rule and the P_∞ average run length are easily obtained, while the corresponding analogue of Page's procedure seems much more difficult to study.

Theorem 3. Let $\mu > 0$ and suppose that in some neighborhood of μ the measure G has positive, continuous density g . Then as $B \rightarrow \infty$

$$E_{0,\mu}(\hat{T}) = \mu^{-2} [2 \log B + \log \log B - 1 - \gamma - \log\{2\pi g^2(\mu)\} - \log(2/\mu^2) + o(1)].$$

The basic idea of the proof of Theorem 3 is already apparent in the arguments of Pollak and Siegmund (1975) or in the more general extension of Lai and Siegmund (1979). A completely rigorous development of the corresponding result in discrete time has been given by Pollak (1983). For the sake of completeness, we give a brief outline in Appendix D.

By comparing Theorem 3 with Propositions 3 and 4, one sees that it is possible asymptotically to do as well to first order as in the case of known $\mu = \delta$. For moderate sample sizes, the higher order terms play an important role, which is investigated numerically in the next section.

5. Numerical Comparisons.

Tables 1 and 2 compare T given by (1) with \tilde{T} given by (2) for the cases $\nu = 0$ and $\nu \rightarrow \infty$, respectively. The values of B and c were chosen so that $E_\infty(T)$ and $E_\infty(\tilde{T})$ are about 790, which seems appropriate for a variety of industrial sampling inspection schemes. The entries in Table 1 were computed by integrating (10) numerically and by applying (14). For Table 2 the asymptotic results of Theorems 1 and 2 were used.

Remark: The numerical integration in (10) can be quite time consuming. For the important range $\mu > \delta/2$ it is possible to show that $E_{0,\mu}(T) = \{\delta(\mu - \delta/2)\}^{-1} \{\log B + \text{const.} + o(1)\}$, so it is possible to do the numerical computation for a moderate value of B and obtain approximations for other B from this one value. Alternatively, in certain ranges one can evaluate the integrals as infinite series to speed up the computations.

Tables 1 and 2 show that the Shirayev-Roberts and Page rules are almost indistinguishable at $\mu = \delta$. For larger values of μ Page's rule does slightly better, while for smaller values the Shirayev-Roberts rule seems preferable. The greatest percentage differences are those favoring Page's rule when μ is large and $\nu = 0$. When $\nu \rightarrow \infty$, the difference favoring Page's rule decreases while the small difference in favor of Shirayev-Roberts remains about what it is for $\nu = 0$.

Since the choice of δ is to some extent arbitrary, it is interesting to observe that in Tables 1 and 2 the choice $\delta = 1/2$ yields much smaller average run lengths for small μ at

Table 1

Comparison of Average Run Lengths		
μ	$E_{0,\mu}(T)$	$E_{0,\mu}(\bar{T})$
$(B = 792, \delta = 1)$		$(c = 6, \delta = 1)$
0	792	793
.25	117	129
.50	34	36
1.0	10.8	10.0
1.5	6.4	5.5
2.0	4.5	3.8
$(B = 791, \delta = .5)$		$(c = 4.65, \delta = .5)$
0	791	791
.25	84	87
.50	32	32
1.0	14.6	11.5
1.5	9.6	7.1
2.0	7.3	5.2

Table 2

Comparison for Large ν and B		
μ	$E_{\nu,\mu}(T - \nu \mid T > \nu)$	$E_{\nu,\mu}(\bar{T} - \nu \mid \bar{T} > \nu)$
$(B = 792, \delta = 1)$		$(\delta^2 B/2 = e^c - 1 - c)$
.25	111	124
.50	31	34
1.0	8.8	9.0
2.0	3.4	3.3
$(B = 791, \delta = .5)$		$(\delta^2 B/2 = e^c - 1 - c)$
.25	71	76
.50	24	25
1.0	9.8	9.4
2.0	4.4	4.1

a relatively minor cost for large μ than does the choice $\delta = 1$. This suggests that one may wish to use a smaller value of δ than the change that one "expects" to occur. In some applications, even the relatively small increase in average run length for large μ entailed by choosing a small δ may be too costly to make this strategy seem reasonable.

Table 3 uses Theorem 3 to give comparable results for \hat{T} defined by (18) with G the distribution of the absolute value of a standard normal random variable. For values of B in the indicated range, the use of a mixture to define \hat{T} seems generally inferior to the practice of using an appropriate, fixed value of δ to define either a Page or a Roberts type rule. The situation changes for larger values of B . Some results are contained in Table 4, which shows that for some inefficiency at $\mu = \delta$, one can do better for extreme μ by using the mixture stopping rule \hat{T} . Since the stopping rule \hat{T} is asymptotically first order optimal simultaneously for all $\mu > 0$, still larger values of B will tend to favor \hat{T} over T . However, this asymptotic optimality seems to take over so slowly as to be irrelevant for many problems.

Table 3
Average Run Lengths for \hat{T} Defined by (18)
with $dG(y) = (2/\pi)^{1/2} \exp(-y^2/2) dy$, $B = 792$

μ	$E_{0,\mu}(\hat{T})$
0	792
.25	133
.50	40
1.0	12.0
1.5	6.1
2.0	4.1

Table 4

Comparisons of \hat{T} and \tilde{T} for large B

μ	$E_{0,\mu}(\hat{T})$	$E_{0,\mu}(\tilde{T})$	$(c = 6.62, \delta = .5)$
	$(B = 5944)$	$(c = 8, \delta = 1)$	
0	5944	5944	5939
.25	202	307	175
.50	57	64	45
1.0	16	14	17
1.5	8.2	7.5	10.3
2.0	5.2	5.1	7.4

6. Open Problems.

We believe that the evidence presented above indicates that there is no persuasive scientific reason for preferring the Page stopping rule to that suggested by Shirayev and Roberts, or vice versa. Depending on the specific context, one or the other might be slightly preferable; but in general the choice may be based essentially on convenience. In this section we indicate some open problems, for which versions of the Shirayev-Roberts rule are more or less obvious and can be studied by techniques similar to those developed here. Page's rule, on the other hand, seems less suited to deal with these new problems. The important distinguishing feature of the Shirayev-Roberts rule is the martingale property utilized in Proposition 1, which seems to have no analogue for cusum procedures.

A particularly interesting article on the applied use of cusum stopping rules is Wilson, et al. (1979). In order to control the quality of radioimmunoassays, plasmas of known composition were occasionally submitted to be assayed, with the result regarded as a normal random variable with unknown mean and variance. The process is regarded as in control if the mean of this random variable is the known composition and if the variance is "small." The stopping rule should detect a change (either increase or decrease) in the mean value or an increase in the variance. Moreover, the target value for the mean is a known quantity; but in the case of the variance the target is whatever can be achieved by careful application of the assaying method - hopefully small - but there is no a priori value which one knows

can be realized.

This problem differs in three important respects from the very simple model described above: (a) two-sided alternatives, (b) a multidimensional parameter space, and (c) an unknown initial parameter value. We shall give here a brief discussion of each of these issues, but defer to a subsequent paper a more thorough investigation.

The issue of two-sided alternatives is the simplest, at least in fairly symmetric problems. The standard modification of Page's procedure is to run two one-sided cusum tests simultaneously, stopping as soon as at least one indicates that a change has occurred. See van Dobben de Bruyn (1968) for a complete discussion. An appropriate modification of the Shirayev-Roberts rule would be to take a mixture of processes as in Section 4, but with the mixing measure G giving positive measure to both positive and negative values of μ . The simplest case would be the measure putting weight 1/2 on $+\delta$ and on $-\delta$. For simple modifications of this sort, comparisons of two procedures yield essentially the same conclusions as in the one-sided case.

For a multidimensional parameter space, the natural generalization of Roberts' rule is again to form a mixture over the parameter space as in Section 4, and the basic theory is much as in the one dimensional case. To obtain a Page like process, it is possible to run simultaneous cusum procedures (Wilson, et al., 1979) or use the method of mixtures (Pollak and Siegmund, 1975); but we have no idea what this does to the average run length under P_∞ .

In many respects the most interesting variation arises when the initial, in control, parameter value is unknown. If the probability model exhibits the appropriate invariance under a group of transformations, as in the case of a normal mean or variance, one can define a Page or Shirayev-Roberts procedure in terms of a maximal invariant function of the data. For example, to detect the change of the drift of Brownian motion from an initial unknown value μ_0 to a new value $\mu_0 + \delta$, the invariant analogue of $R(t)$ defined in (3) is

$$R^*(t) = \int_0^t \exp[\delta\{sW(t)/t - W(s)\} - \delta^2 s(1-s/t)/2] ds.$$

Relative to the appropriate σ -fields the process $R^*(t) - t$ is a P_∞ martingale. Hence if

$T^* = \inf\{t : R^*(t) = B\}$, exactly as in Proposition 1 we have that $E_\infty(T^*) = B$. Although an analogous Page type procedure is easily defined, we have no idea what its P_∞ average run length is.

An entirely new feature of the problem of the preceding paragraph is that evaluation of $E_\nu(T^* - \nu \mid T^* > \nu)$ only at the extreme values of $\nu = 0$ and $\nu \rightarrow \infty$ is uninteresting. If a "change" occurs at $\nu = 0$, it cannot be detected because the new value of μ cannot be distinguished from the initial value. If the change takes place after an extremely long period of time, we have so much data to estimate μ that we are effectively back in the situation where μ is known. We expect to discuss this model in a subsequent paper.

Appendix A

We sketch here the optimality considerations of Shirayev (1963) and Pollak (1984).

Assume that μ equals either 0 or δ .

Assuming that ν has an exponential distribution with mean $1/\lambda$ and that if one stops at t , the loss is 1 or $(t - \nu)$ according as $t < \nu$ or $t \geq \nu$, Shirayev (1963) shows that the Bayes rule stops at

$$\tau(\lambda, c) = \inf\{t : P\{\nu < t \mid W(s), s \leq t\} \geq b(\lambda, c)\}.$$

From Shirayev's formula for $b(\lambda, c)$ it follows that $b_c = \lim_{\lambda \rightarrow 0} b(\lambda, c)/\lambda$ exists and satisfies $c \exp(\delta/b_c) \int_{\delta/b_c}^{\infty} y^{-1} \exp(-y) dy = 1$. Hence T_B defined by (1) is a limit of Bayes rules for a particular $c = c_B$.

It is possible to modify T_B slightly to make its risk essentially constant in ν and hence to make T_B approximately minimax. Let $R'(t)$ denote the process $R(t)$ started off from the P_∞ -limiting distribution of $R(s)$ given $T_B > s$ as $s \rightarrow \infty$. Let $T'_B = \inf\{t : R'(t) \geq B\}$. Then as $B \rightarrow \infty$

$$\begin{aligned} \sup_{\nu} E_{\nu}(T'_B - \nu \mid T'_B > \nu) &= \inf_{\tau} \sup_{\nu} E_{\nu}(\tau - \nu \mid \tau > \nu) + o(1) \\ &= 2\delta^{-2} \{\log(\delta^2 B/2) - 1 - \gamma\} + o(1), \end{aligned}$$

where the inf is over all rules τ with $E_{\infty}(\tau) \geq E_{\infty}(T'_B)$. The first equality follows from arguments similar to but much easier than those of Pollak (1984), Theorem 2. The second is a consequence of Theorem 1 of this paper.

Appendix B

Here we discuss calculation of the integrals involved in the proof of Theorem 1. It suffices to evaluate (up to terms converging to 0 as $B \rightarrow \infty$)

$$E_{0,\mu}^0(T_B) = \int_0^B E_{0,\mu}^0(T_x) d\bar{M}_0(x), \quad (20)$$

where \bar{M}_0 denotes the speed measure M_0 normalized to be a probability (this is the P_∞ limiting distribution of $R(t)$), and by (10) $E_{0,\mu}^0(T_x) = 2 \int_0^x \int_y^x dS_\mu(z) dM_\mu(y)$. Inverting the order of integration in (20), so that we integrate first with respect to M_μ , then M_μ , yields

$$\int_0^B E_{0,\mu}^0(T_x) d\bar{M}_0(x) = 2 \int_0^B \{ \exp(-2/\delta^2 B) - \exp(-2/\delta^2 z) \} \int_0^z dM_\mu(y) dS_\mu(z).$$

The first term on the right hand side can be recognized to equal $E_{0,\mu}^0(T_B) \exp(-2/\delta^2 B)$, while the second, after another inversion in the order of integration, becomes

$$2\delta^{-2} \int_0^B y^{2(\mu/\delta-1)} \exp(-2/\delta^2 y) \int_y^B z^{-2\mu/\delta} dz dy. \quad (21)$$

Putting these facts together shows that (20) equals the sum of (21) and

$$E_{0,\mu}^0(T_B) \{ 1 - \exp(-2/\delta^2 B) \}. \quad (22)$$

As $B \rightarrow \infty$, for any $\mu > 0$ $E_{0,\mu}^0(T_B) = o(B)$, so (22) is asymptotically negligible. Hence it remains to evaluate (21), which is a tedious but fairly straightforward job and yields the results given in Theorem 1.

Appendix C

Shiryayev (1963) discusses "detection of destruction of a stationary regime," which superficially resembles our Section 3. Here we attempt to indicate some important differences.

In addition to our basic assumptions, Shiryayev assumes (i) when the stopping rule (1) or (2) indicates a change, we can immediately ascertain whether a change has indeed occurred, restarting the process (from 0) if there has been no change; and (ii) this new process (i.e. the original process, renewed at each false alarm) has been running for an extremely long time without any changepoint, so that the number of false alarms already observed is becoming infinitely large. Mathematically this means that $\nu \rightarrow \infty$, perhaps $B \rightarrow \infty$; but in any case $\nu/B \rightarrow \infty$.

Shiryayev's formulation is presumably reasonable if the cost associated with a false alarm is relatively small compared to the cost of observation after the changepoint ν . On the other hand, we envision situations where the cost of a false alarm is substantial, and/or it is difficult to tell immediately whether an alarm is a false one. Thus we have taken ν and B simultaneously large with no assumed relation between them.

With some reformulation, it is possible to bring about a partial unification of the two viewpoints.

Suppose with Shiryayev that at each false alarm the detection process is immediately restarted from scratch. Let S_1, S_2, \dots denote the times of successive alarms, and let $L = L(t) = \max\{k : S_k \leq t\}$ be the number of alarms before time t . Then $L(t)$ is a P_∞ -renewal process with renewal epochs S_n . Suppose that we agree to measure delay in detecting a disorder not by $E_\nu(S_1 - \nu \mid S_1 > \nu)$ as we have in the rest of this paper, but by

$$E_\nu(S_{L(\nu)+1} - \nu) = \int_0^\nu E_t(S_1 - t \mid S > t) P_\nu\{\nu - S_{L(\nu)} \in dt\}. \quad (23)$$

Assuming $\mu = \delta$, Shiryayev uses renewal theory to evaluate (23) as $\nu \rightarrow \infty$ and then evaluates this limit as $B \rightarrow \infty$. Assume now that ν and B simultaneously become infinitely large, in any relation whatever. For either of the detection rules (1) or (2), it is easy to use the results of Theorems 1 and 2, together with the fact that the P_∞ distribution of $\nu - S_{L(\nu)}$

becomes more diffuse as $E_\infty(S_1) \rightarrow \infty$, to show that the asymptotic evaluations given in Theorems 1 and 2 are also satisfied by the new criterion (23) for any $\mu > 0$.

Appendix D

In this appendix we indicate very informally the ideas leading to Theorem 3. We assume throughout that $\nu = 0$ and write E_μ to denote expectation.

For the stopping rule (1) with $\mu = \delta$, we have

$$\log B = \mu W(T) - \mu^2 T/2 + \log \int_0^T \exp\{-\mu W(s) + \mu^2 s/2\} ds,$$

so by Wald's identity

$$\log B = \frac{1}{2} \mu^2 E_\mu(T) + E_\mu \left[\log \int_0^T \exp\{-\mu W(s) + \mu^2 s/2\} ds \right]. \quad (24)$$

As $B \rightarrow \infty$, $T = T_B \rightarrow \infty$ in probability, so

$$E_\mu \left[\log \int_0^T \exp\{-\mu W(s) + \mu^2 s/2\} ds \right] = E_\mu \left[\log \int_0^\infty \exp\{-\mu W(s) + \mu^2 s/2\} ds \right] + o(1), \quad (25)$$

and this last expectation can be evaluated exactly by consideration of (12), (24), and (25).

For the stopping rule (18), some algebra yields

$$\begin{aligned} \log B &= \mu W(\hat{T}) - \mu^2 \hat{T}/2 - \frac{1}{2} \{W(\hat{T}) - \mu \hat{T}\}^2 / \hat{T} - \frac{1}{2} \log \{\hat{T}/(2\pi)\} \\ &\quad + \log \int_0^{\hat{T}} \hat{T}^{1/2} \varphi[\hat{T}^{1/2} \{y - W(\hat{T})/\hat{T}\}] \int_0^\infty \exp\{-y W(s) + y^2 s/2\} ds dG(y) \end{aligned} \quad (26)$$

The (random) measure $d\hat{G}(y) = \hat{T}^{1/2} \varphi[\hat{T}^{1/2} \{y - W(\hat{T})/\hat{T}\}] dG(y)$ becomes progressively more concentrated around $W(\hat{T})/\hat{T}$ as B and hence $\hat{T} \rightarrow \infty$, and with overwhelming P_μ -probability, $W(\hat{T})/\hat{T}$ becomes concentrated around μ . Hence as $B \rightarrow \infty$ the expectation of the final expression in (26) converges to the expectation appearing on the right hand side of (25), which can be evaluated. By Wald's identity and some additional calculations we have as $B \rightarrow \infty$

$$\begin{aligned} E_\mu \{\mu W(\hat{T}) - \mu^2 \hat{T}/2\} &= \frac{1}{2} \mu^2 E_\mu(\hat{T}), \\ E_\mu \{ \{W(\hat{T}) - \mu \hat{T}\}^2 / \hat{T} \} &\rightarrow 1, \end{aligned}$$

and

$$E_\mu \{\log(\hat{T})\} \sim \log\{E_\mu(\hat{T})\} \sim \log\{(2 \log B)/\mu^2\}.$$

Substituting these results into (26) yields Theorem 3.

References

van Dooben de Bruyn, C. S. (1968). *Cumulative Sum Tests*, London, Griffin.

Karlin, S. and Taylor, H. (1981). *A Second Course in Stochastic Processes*, Academic Press, New York.

Lai, T. L. and Siegmund, D. (1979). A non-linear renewal theory with applications to sequential analysis II, *Ann. Statist.* **7**, 60-76.

Loève, M. (1963). *Probability Theory*, 3rd ed., Van Nostrand, Princeton.

Lorden, G. (1971). Procedures for reacting to a change in distribution, *Ann. Math. Statist.* **42**, 1897-1908.

Page, E. S. (1954). Continuous inspection schemes, *Biometrika* **41**, 100-115.

Pollak, M. (1984). Optimal detection of a change in distribution, to appear in *Ann. Statist.*

Pollak, M. (1983). Average run lengths of an optimal method of detecting a change in distribution, to appear in *Ann. Statist.*

Pollak, M. and Siegmund, D. (1975). Approximations to the expected sample size of certain sequential tests, *Ann. Statist.* **3**, 1267-1281.

Roberts, S. W. (1966). A comparison of some control chart procedures, *Technometrics* **8**, 411-430.

Shiryayev, A. N. (1963). On optimal methods in earliest detection problems, *Theory of Probab. and Its Applications* **8**, 26-51.

Weatherall, J. A. C. and Haskey, J. C. (1976). Surveillance of malformations, *Br. Med. Bull.* **32**, 39-44.

Wilson, D. W., Griffiths, K. Kemp, K. W., Nix, A. B. J., Rowlands, R. J. (1979). Internal quality control of radioimmunoassays: monitoring of error, *Jour. Endocr.* **80**, 365-372.

END

FILMED

3-85

DTIC